Lincoln Technologies, Inc. and IBM
System Test Report

System Title:	**WebSDM/JANUS Load Test for JANUS Phase 2 Deployment**		
Client Organization:	**NCI**		
Client Project Manager:	**Christo Andonyadis**		
IBM Project Manager:	**Mary Ferramosca**		
Lincoln Project Manager:	**Sally Cassells**		
Document File Name:	**LTI and IBM Janus System Load Test Report v1.1.doc**	**Version:**	**1.1**
Prepared by:	**IBM/LTI team**	**Date:**	**10/03/07**
Document Status:	**Final**		

Version History

Version	Date	Author	Comments
0.1	09-September-2007	Mary Ferramosca	Initial Version
0.2	September 10, 2007	Luting Li (Olivia)	Added - 1) Test Summary; 2) Comments and Recommendations
0.3	September 11, 2007	Sally Cassells	Data preparation section addition and editorial changes
0.4	September 11, 2007	Luting Li (Olivia)	Section 3 changes
0.5	September 13,2007	Mary Ferramosca	Editorial changes
0.6	September 14, 2007	Sally Cassells	Added data flow diagram. Updated section 2.2
0.7	September 14, 2007	Houtan Aghili	Comments made
0.8	September 14, 2007	Luting Li (Olivia)	Added performance metrics
0.9	September 14, 2007	Sally Cassells	Added WebSDM load performance information
1.0	September 26, 2007	Luting Li (Olivia)	Clarified section 3 - item 2 to address Jay Levin's comment from FDA
1.1	October 03,2007	Mary Ferramosca	Added Christo and Bobbie's acceptance dates to the Review/Acceptance Table.

Review/Acceptance Table

Person Reviewing/ Approving	Date	Comments
Sally Cassells – Project Manager (LTI)	14-September-2007	
Mary Ferramosca – Project Manager (IBM)	14-September-2007	
Bobbie Witczak - (FDA)	03-October-2007	
Christo Andonyadis (NCI)	03-October-2007	

Table of Contents

1. Introduction

1.1 Purpose

The purpose of this System Test Report is to document the results of the Large Study Load testing performed by the IBM/LTI team. This work is part of the JANUS Phase 2 Deployment project funded by the NCI.

1.2 Scope

The scope of this System Test Report is a description of the results of loading a large sample study into the WebSDM staging area and then into the JANUS data warehouse in the NCI QA environment. The report includes descriptions of deviations, assumptions, exclusions, limitations, and risks.

2. Test Summary

2.1 Test Set-up

The test was performed in the NCI QA environment. The following integrated components were deployed to QA environment and tested:

- JANUS staging area
  - WebSDM Integration Server (WIS) – Open Source stand alone module
  - WebSDM (Commercial Off the Shelf software)
- JANUS
  - JANUS Web Services
  - JANUS Data Load Package

2.2 Test Data Generation

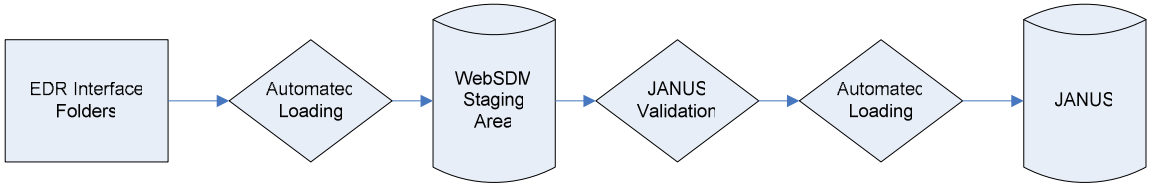
The test data used for this activity was adapted from sponsor data provided to LTI in a past customer engagement. The data includes 12 datasets with 8556 subjects.

To de-identify the data and ensure that it could be successfully loaded into JANUS, the following set of data transformations was performed:

- Subject identifiers (USUBJID, SUBJID, STUDYID, SITEID) were set to automatic sequence values and propagated to all domains.
- Data was cleaned/modified to ensure no 'Fail-reject' Janus Validation errors:
  - Invalid ISO8601 date values were re-formatted (all domains)
  - Codelist definitions in the Define.XML file were modified to include all coded values.
  - Records where xxENDTC < xxSTDTC were removed.
  - The SAS missing value character (.) was replaced with NULL where found in data fields
  - Values of AEDECOD were mapped to MedDRA 6.0 Preferred Terms (from at LLT vlaues)
  - Rows where the domain Topic Variable value was missing were removed
  - For missing SDTM Required fields, hard coded values were 'derived'
- New Define.XML was generated.

2.3 Test Data Flow

A complete description of the JANUS data loading process is available in the Detailed Design Specification for the EDR/WebSDM/JANUS Interface (available on G-FORGE). A schematic depiction of the process is provided in the diagram below.



The sample study input data is in the form of SAS Transport files and includes the domain datasets listed below:

Dataset File Name	Number of Data Rows
AE.xpt	27257
DM.xpt	8556
DS.xpt	8932
EX.xpt	8556
LB.xpt	355060
SC.xpt	42657
SE.xpt	28521
SU.xpt	59892
TA.xpt	3
TE.xpt	3
TS.xpt	23
TV.xpt	14

2.4 Test Result

The following table lists the test results grouped by submitted SDTM domain.

Table 1 - Test Result

Input		Data Load Result		Pass /Fail	Note
SDTM Domain	**Number of Records**	**JANUS Tables**	**Number of Records Created**		
DM	8556	SUBJECTS	8556	P	
TA	3	ARMS	1	P	Only 1 distinct armcd/arm
TA	3	PLAN_ELEMENT_SEQUENCE	3	P	
TE	3	ELEMENTS	3	P	
TV	14	PALNNED_VISITS	14	P	
SE	28521	ELEMENT_SEQUENCES	28521	P	
LB	355060	FINDINGS - LB	355060	P	
SC	42657	FINDINGS - SC	42657	P	
AE	27257	EVENTS - AE	27257	P	
DS	8932	EVENTS - DS	8932	P	
EX	8556	INTERVENTIONS - EX	8556	P	
SU	59892	INTERVENTIONS - EX	59892	P	
TS	23	COMMENTS	23	P	There are 23 parent rows in COMMENTS and each corresponds to a TS record.

2.5 Performance Metrics

The below table listed timing information of loading the large study, in comparison with the other two sample studies (LTI sample study and IBM Taxotere/Iressa sample study) that were previously loaded in QA environment for milestone 3.

Study	**Study Profile**	**Time** **(Loading SDTM Datasets from SAS Transport Files into WebSDM Staging Tables)**	**Time** **(Loading SDTM Datasets from WebSDM Staging Tables into JANUS Database)**	**Time** **(Total)**
Large Study	AE 27257 DM 8556 DS 8932 EX 8556 LB 355060 SC 42657 SE 28521 SU 59892 TA 3 TE 3 TS 23 TV 14	1 hour 30 min	16 hours 40 min	18 hours 10 min
LTI Sample Study	TA 16 TE 5 TV 10 DM 72 SE 269 SV 715 LB 2254 PE 1584 SC 200 VS 10752 DA 20	4 min	7 min	11 min

Study	Study Profile	Time (Loading SDTM Datasets from SAS Transport Files into WebSDM Staging Tables)	Time (Loading SDTM Datasets from WebSDM Staging Tables into JANUS Database)	Time (Total)
	AE 333 DS 134 MH 736 EX 94 CM 161 IE 5 TI 35 CO 22 SQ 96 RELREC 4 TS 4			
IBM Taxotere/Iressa Sample Study	TA 19 TE 5 TV 68 DM 316 SE 2395 SV 8052 LB 42125 PE 10987 QS 62016 SC 1546 VS 39708 ON 5804 AE 277 DS 948 MH 316 EX 732 SU 316 CM 271	10 min	2 hours 30 min	2 hours 40 min

Study	Study Profile	Time (Loading SDTM Datasets from SAS Transport Files into WebSDM Staging Tables)	Time (Loading SDTM Datasets from WebSDM Staging Tables into JANUS Database)	Time (Total)
	IE 2 TI 17 CO 4876 SQ 93			

2.6 Test Limitations, Exclusion and Associated Risk

Only one study was tested. Although the original data was from a planned FDA submission and therefore a realistic test case, it does not necessarily reflect a widely used study structure or represent a standard study organization. With 8556 subjects it is probably larger than the 'average' study received by the FDA.

3. Comments and Recommendations

Two issues were discovered during the test that warrants further discussion and process improvement.

1) Subject Visit

   When the SV domain is not submitted, the JANUS loading program code provides a workaround to derive the SV information it needs. The Visit Number (VISITNUM), Visit Name (VISIT) and Planned Study Day of the Visit (VISITDY) variable values are extracted from Findings, Events and Interventions domains provided and inserted into the JANUS VISITS table.

   If use of the visit identification variables (VISTNUM, VISIT and VISITDY) in Findings, Interventions and Events for the same subject is not consistent, information in JANUS VISITS table will be redundant and low quality. As an example, in the test data, where no SV domain was provided, the VISITS table includes such cases shown in the table below:

subject_id	NUM (VISITNUM)	NAME (VISIT)
123	1	
123	1	BASE

In addition, when there is a large volume of records in the domains of Observation Class, the extraction of the visit information and counting the number of the visits records embedded in these record to create statistics has negative impact on the data load performance.

The workaround solution could be modified by adding additional transformation logic to consolidate multiple records for the same visit, based on visitnum or visit, but the transformation logic will be complex and further degrade the load performance.

A better solution is to promote the requirement of SV domain submission or complete and consistent submission of the available visit information across the submitted domains of Observation Class if SV domain is not submitted.

2) SDTM Variable XXORRES, XXSTRESN and XXSTRESC Mapping

SDTM variable XXORRES (result or finding in original unit) was originally mapped to either JANUS table FINDINGS column continuous_value if it has numeric value or table COMMENTS column if it has non-numeric value. The screening logic was based on whether it only contains '.', '+', '-', '0-9'.

The screening logic was found broken when the test data XXORRES has value like '3-5'. Although additional screening logic can be added for this type of value pattern, it is hard to foresee all possible value patterns. Due to the time constraint, a quick and safe fix by storing XXORRES in COMMENTS.comm was implemented for this phase. The change implemented has no impact on the data analysis through tools (I-Review/J-Review, WebSDM) because the study data in JANUS are exposed to the tools via SDTM materialized views and the original results in the views will be expressed the same way as they were submitted. However, the long term fix is to change JANUS data model by adding a respective field in the table FINDINGS to

store the value of XXORRES. A change request has to be submitted to JANUS Change Control Board to get approval before implementing the change.

SDTM variable XXSTRESN (numeric result or finding in standard unit) is currently mapped to table COMMENTS column comm. Because the SDTM Implementation Guide defines XXSTRESN with Number data type, this mapping should be changed. XXSTRESN should be mapped to table FINDINGS column continuous_value. This change will impact the current mapping of some other related SDTM variables as well as the script of creating materialized views. This change will be considered for implementation in the next phase.

SDTM variable XXSTRESC (character result or finding in standard format) is currently mapped to table COMMENTS column comm. A better way to accommodate this variable is to make change in JANUS data model by adding a respective field in table FINDINGS. A change request has to be submitted to JANUS Change Control Board to get approval before implementing the change.

4. External References

Document	Accessible At
Study Data Tabulation Model Dataset Validation Requirements	JANUS G-Forge Website
JANUS Web Services Design Document	JANUS G-Forge Website
CDISC Study Data Transmission Model Implementation Guide Version 3.1.1	http://www.cdisc.org/models/sdtm/v1.1/index.html